P-2212

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-231238
(43)Date of publication of application : 05.09.1997

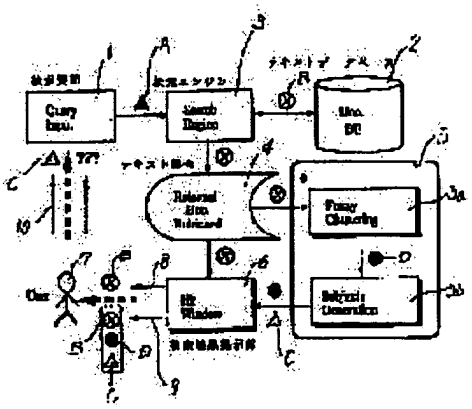
(51)Int.Cl. G06F 17/30

(21)Application number : 08-058391
(22)Date of filing : 20.02.1996

(71)Applicant : OMRON CORP
(72)Inventor : GO ATOU
SOGO TAIJI
SAWADA AKIRA

(54) DISPLAY METHOD FOR TEXT RETRIEVAL RESULT AND DEVICE THEREFOR

(57)Abstract:
PROBLEM TO BE SOLVED: To improve both retrieval efficiency and accuracy by dividing a text set into plural groups based on the theme analysis result of every text, generates the theme sort information showing the attribute of every group, and displays these information in every group.
SOLUTION: A retrieval engine 3 expands a retrieval expression based on a prescribed algorithm and extracts a relative text set 4 out of a document data base 2. A fuzzy gathering part 5a of a processing part 5 divides the set 4 into plural groups based on the theme analysis result of every text, and a theme sort information generation part 5b generates the theme sort information showing the attribute of every group. A retrieval result display part 6 processes the acquired information (text set B, centroid D and theme sort information C) in a prescribed display mode and shows them to a user 7. As a result, the document retrieval result can be easily confirmed and the retrieval efficiency and accuracy can be improved owing to prevention of the retrieval omission.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平9－231238

(43)公開日　平成9年(1997)9月5日

(51)Int.Cl.⁶	識別記号	庁内整理番号	FI		技術表示箇所
G06F 17/30			G06F 15/403	370A	
			15/401	310D	
			15/403	380E	

審査請求　未請求　請求項の数24　FD　（全 19 頁）

(21)出願番号　特願平8－58391

(22)出願日　平成8年(1996)2月20日

(71)出願人　000002945
オムロン株式会社
京都府京都市右京区花園土堂町10番地

(72)発明者　呉　亜輝
京都府京都市右京区花園土堂町10番地　オムロン株式会社内

(72)発明者　十河　太治
京都府京都市右京区花園土堂町10番地　オムロン株式会社内

(72)発明者　澤田　晃
京都府京都市右京区花園土堂町10番地　オムロン株式会社内

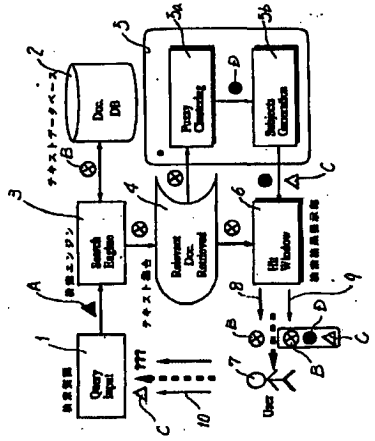
(74)代理人　弁理士　飯塚　信市

(54)【発明の名称】　テキスト検索結果表示方法及び装置

(57)【要約】

【課題】　文書検索結果に対する確認を容易として、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができ、しかも、提示された主題情報がデータを如何に効率的に絞り込めるかの指針にもなり、この付加された応答情報を利用して高度な適応検索（Relevance Feedback）を行い得る。

【解決手段】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割し、該分割により得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成し、該生成された各グループの主題分類情報をグループ別に区分して表示する。



【特許請求の範囲】

【請求項１】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップと、
前記分割ステップによって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成ステップと、
前記生成ステップで求めた各グループの主題分類情報をグループ別に区分して表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。

【請求項２】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップと、
前記分割ステップによって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成ステップと、
前記各グループのそれぞれについて、そのグループと前記検索条件との間の適合度を求めるグループ適合度算出ステップと、
前記生成ステップで求めた各グループの主題分析情報を、前記適合度算出ステップによって求めた適合度の大きい順に、グループ別に区分して表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。

【請求項３】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップと、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出ステップと、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択ステップと、
前記選択ステップで選択されたグループ内のテキストを前記算出された所属度の順に内容表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。

【請求項４】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップと、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップと、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択ステップと、
前記選択ステップで選択されたグループ内のテキストを前記算出された適合度の順に内容表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。

【請求項５】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップと、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出ステップと、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップと、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための表示対象グループ選択ステップと、
前記各グループ内のテキストを検索条件への適合度順に表示するか、或いは当該グループへの所属度の順に表示するかを選択するための表示順序基準選択手段と、
前記表示対象グループ選択ステップで選択されたグループ内のテキストを前記表示順序基準選択手段にて選択された表示順序基準の順に内容表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。

【請求項６】　前記分割ステップは、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を、ファジイ・クラスタリング法を用いて複数個のグループに分割する、
ことを特徴とする請求項１乃至請求項５のいずれかに記載のテキスト検索結果表示方法。

【請求項７】　前記生成ステップにて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を幾つかのキーワードの組により表すものである、
ことを特徴とする請求項１若しくは請求項２のいずれかに記載のテキスト検索結果表示方法。

【請求項８】　前記生成ステップにて生成される当該グループの属性を表現する主題分類情報は、当該部ループの属性を短い文章により表すものである、
ことを特徴とする請求項１若しくは請求項２のいずれかに記載のテキスト検索結果表示方法。

【請求項９】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合の特徴行列に対してファジイ・クラスタリングを行い、各文　毎に各分類カテゴリーへの所属度を生成する所属度生成ステップと、
前記生成された所属度を用いて、各文書を１若しくは２以上の分類カテゴリーに割り付ける文書割り付けステップと、
前記複数個の分類カテゴリーの中で、テキスト表示対象

となる分類カテゴリーを選択するための分類カテゴリー選択ステップと、
前記分類カテゴリー選択ステップで選択された分類カテゴリー内のテキストをそのグループに対する適合度の順に内容表示する表示ステップとを具備する、
ことを特徴とするテキスト検索結果表示方法。
【請求項１０】　前記文書割り付けステップは、各文書をその所属度の上位ｋ個の分類カテゴリーに割り付ける、
ことを特徴とする請求項９に記載のテキスト検索結果表示方法。
【請求項１１】　前記文書割り付けステップは、各文書をある閾値α以上の所属度値を有する分類カテゴリーに割り付ける、
ことを特徴とする請求項９に記載のテキスト検索結果表示方法。
【請求項１２】　前記文書割り付けステップは、各文書をカテゴリーの確率分布を考慮して分類カテゴリーに割り付ける、
ことを特徴とする請求項９に記載のテキスト検索結果表示方法。
【請求項１３】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段と、
前記分割手段によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成手段と、
前記生成手段で求めた各グループの主題分類情報をグループ別に区分して表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項１４】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段と、
前記分割手段によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成手段と、
前記各グループのそれぞれについて、そのグループと前記検索条件との間の適合度を求めるグループ適合度算出手段と、
前記生成手段で求めた各グループの主題分析情報を、前記適合度算出手段によって求めた適合度の大きい順に、グループ別に区分して表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項１５】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段と、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出手段と、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択手段と、
前記選択手段で選択されたグループ内のテキストを前記算出された所属度の順に内容表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項１６】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段と、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出手段と、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択手段と、
前記選択手段で選択されたグループ内のテキストを前記算出された適合度の順に内容表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項１７】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段と、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出手段と、
前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出手段と、
前記複数個のグループの中で、テキスト表示対象となるグループを選択するための表示対象グループ選択手段と、
前記各グループ内のテキストを検索条件への適合度順に表示するか、或いは当該グループへの所属度の順に表示するかを選択するための表示順序基準選択手段と、
前記表示対象グループ選択手段で選択されたグループ内のテキストを前記表示順序基準選択手段にて選択された表示順序基準の順に内容表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項１８】　前記分割手段は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を、ファジイ・クラスタリング法を用いて複数個のグループに分割する、
ことを特徴とする請求項１３乃至請求項１７のいずれかに記載のテキスト検索結果表示装置。
【請求項１９】　前記生成手段にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を幾つかのキーワードの組により表すものである、
ことを特徴とする請求項１３若しくは請求項１４のいずれかに記載のテキスト検索結果表示装置。
【請求項２０】　前記生成手段にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を短い文章により表すものである、
ことを特徴とする請求項１３若しくは請求項１４のいずれかに記載のテキスト検索結果表示装置。
【請求項２１】　与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合の特徴行列に対してファジイ・クラスタリングを行い、各文書毎に各分類カテゴリーへの所属度を生成する所属度生成手段と、
前記生成された所属度を用いて、各文書を１若しくは２以上の分類カテゴリーに割り付ける文書割り付け手段と、
前記複数個の分類カテゴリーの中で、テキスト表示対象となる分類カテゴリーを選択するための分類カテゴリー選択手段と、
前記分類カテゴリー選択手段で選択された分類カテゴリー内のテキストをそのグループに対する適合度の順に内容表示する表示手段とを具備する、
ことを特徴とするテキスト検索結果表示装置。
【請求項２２】　前記文書割り付け手段は、各文書をその所属度の上位ｋ個の分類カテゴリーに割り付ける、
ことを特徴とする請求項２１に記載のテキスト検索結果表示装置。
【請求項２３】　前記文書割り付け手段は、各文書をある閾値α以上の所属度値を有する分類カテゴリーに割り付ける、
ことを特徴とする請求項２１に記載のテキスト検索結果表示装置。
【請求項２４】　前記文書割り付け手段は、各文書をカテゴリーの確率分布を考慮して分類カテゴリーに割り付ける、
ことを特徴とする請求項２１に記載のテキスト検索結果表示装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】この発明は、文書データベースの検索に好適なテキスト検索結果表示方法及び装置に係り、特に、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割し、該分割により得られた各グループの属性を表現する主題分類情報をグループ別に区分して表示するようにしたテキスト検索結果表示方法及び装置に関する。
【０００２】
【従来の技術】従来のテキスト検索結果表示装置としては、例えば、特開平６－７６００４号公報に記載されたものが知られている。

【０００３】この装置は、データベース検索結果を格納するデータベース検索解格納部と、前記データベース検索解が有する複数の属性値に利用者の制御入力を加味して各検索解間の距離を算出する検索解間距離算出手段と、検索解間距離を用いて検索解を利用者に指定した個数或いは予め定められた個数のグループに分割する検索解グループ分割手段と、所属グループの重心付近に位置する検索解を算出するグループ代表検索解算出手段と、各グループの代表検索解の中から特定の検索解を利用に選択させる代表検索解選択手段と、代表検索解が属しているグループ内の全検索解を表示するグループ内検索解表示手段とから構成されている。
【０００４】すなわち、この従来装置にあっては、non-overlapping手法で構造化された（数値）データベース検索解をユーザーの指定した分類数に分類するもので、分類されたグループの重心に最も近い検索解を１件ずつパイロットデータとして表示して利用者に希望するグループを選択させ（順位付けなし）、選択されたグループ内の全検索解をランキングせずに表示するものである。
【０００５】
【発明が解決しようとする課題】しかしながら、このような従来のテキスト検索結果表示装置（検索解表示装置）にあっては、次の理由により、フルテキストのような非構造化データベースへの適用は困難であると言う問題点があった。
【０００６】すなわち、このような従来装置にあっては、グループ内の重心位置の代表検索解が表示されるため、代表検索解がグループ内の代表文　である場合には、その代表文書の内容を端的に表すものが表示されず、文書全体が表示されるのでグループの内容が把握し難い。つまり、分類された各グループの主題意味を提示するために、単なるグループの重心に最も近い検索解を１件ずつパイロットデータとして表示するだけでは、内容的に特定すぎる場合があり、むしろ、グループ内の共通的な属性項目群を抽出し、利用　に提示することが好ましい。加えて、フルテキスト検索システムの場合にあっては、パイロットデータとして全ての属性データをそのまま提示することは無意味であり、文書内容を容易に理解できるようなパイロットデータの新しい定義が望まれる。
【０００７】また、従来装置にあっては、グループが検索条件に対する適合度の順に並べられないので、検索目的に合致したグループを選択し難い。加えて、従来装置にあっては、グループ内の解がグループへの所属度の順に並んでいないので、グループの代表解を参照するだけでは、グループのイメージが把握し難い場合でも、他の解を参照してイメージを把握することが困難である。つまり、選択されたグループ内の全検索解をランキングせずに表示する方式では、分類件数が多くなると、検索結

果への特定のために利用者の負担が大きくなる。このような負担を軽減して検索効率を向上させるためには、検索結果への特定を促進できるようなランキング機能が望まれる。

【０００８】更に、文書は複数の主題を持っているのが通例であるため、一つの文書を一つのクラスタにしか分類できない従来の手法では、文書分類結果の表示上では検索結果に漏れを生じる虞れがある。そのため、文書検索結果集合に対し主題分類を行う際に複数の異なる（主題を表す）クラスタに属することを許すようなoverlapping手法が望まれる。

【０００９】この発明は、上述の問題点に鑑みてなされたものであり、その目的とするところは、文書検索結果に対する確認を容易として、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができ、しかも、提示された主題情報がデータを如何に効率的に絞り込めるかの指針にもなり、この付加された応答情報を利用して高度な適応検索（Relevance Feedback）を行い得るようにした検索結果表示方法及び装置を提供することにある。

【００１０】

【課題を解決するための手段】この出願の請求項１（又は請求項１３）に記載の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップ（又は手段）と、前記分割ステップ（又は手段）によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成ステップ（又は手段）と、前記生成ステップ（又は手段）で求めた各グループの主題分類情報をグループ別に区分して表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００１１】ここで、『データベース』とは、ハードディスクや光ディスク等の大容量記憶媒体に記憶されたテキスト集合やインターネット上に存在するホームページ等のテキスト集合がこれに相当する。

【００１２】また、『主題分析』とは、テキストの内容を端的に示す情報を生成することを意味するものであり、文書内のタイトル上のキーワードの集合を生成するものであっても良い。実施の形態においては、文書を文書空間での特徴ベクトルで表現しているベクトル（Ｆｉ）がこれに相当する。

【００１３】また、『主題分類情報』とは、テキストのグループについて、そのグループの内容を端的に示す情報を意味する。実施の形態では、キーワード方式とテキスト方式との２方式が示されている。

【００１４】そして、この請求項１（又は請求項１３）の発明によれば、グループを端的に表現する情報を付加してグループ別に区分表示するので、検索結果を構成するグループの全体像を把握し易くなり、次の処理のためのグループ選択が非常に容易となる。

【００１５】この出願の請求項２（又は請求項１４）の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップ（又は手段）と、前記分割ステップ（又は手段）によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成ステップ（又は手段）と、前記各グループのそれぞれについて、そのグループと前記検索条件との間の適合度を求めるグループ適合度算出ステップ（又は手段）と、前記生成ステップ（又は手段）で求めた各グループの主題分析情報を、前記適合度算出ステップによって求めた適合度の大きい順に、グループ別に区分して表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００１６】そして、この請求項２（又は請求項１４）の発明によれば、前記請求項１（又は請求項１３）に記載の発明の効果に加えて、検索条件への適合度の順に表示するので、検索目的に合致したグループをグループの内容を確認しながら選択することができる。

【００１７】この出願の請求項３（又は請求項１５）の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出ステップ（又は手段）と、前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択ステップ（又は手段）と、前記選択ステップ（又は手段）で選択されたグループ内のテキストを前記算出された所属度の順に内容表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００１８】そして、この請求項３（又は請求項１５）の発明によれば、選択されたグループ内のテキストがグループへの所属度の順に表示されるので、グループの定義が把握し易くなる。

【００１９】この出願の請求項４（又は請求項１６）の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップ（又は手段）と、前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択ステップ（又は手段）と、前記選択ステップ（又は手段）で選択されたグループ内のテキストを前記算出された適合度の順に内容表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００２０】そして、この請求項４（又は請求項１６）の発明によれば、検索条件に適したグループを選択し、さらにその中のテキストを検索条件の順に表示するので、検索結果をグループ分けしないでテキストを適合度順に表示する場合よりも、検索条件に対して適切なテキストが早く確実に表示される。

【００２１】この出願の請求項５（又は請求項１７）の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップ（又は手段）と、前記複数個のグループの中で、テキスト表示対象となるグループを選択するための表示対象グループ選択ステップ（又は手段）と、前記各グループ内のテキストを検索条件への適合度順に表示するか、或いは当該グループへの所属度の順に表示するかを選択するための表示順序基準選択ステップ（又は手段）と、前記表示対象グループ選択ステップで選択されたグループ内のテキストを前記表示順序基準選択手段にて選択された表示順序基準の順に内容表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００２２】そして、この請求項５（又は請求項１７）の発明によれば、ユーザーの目的に応じてテキストの表示順序を変えることができる。

【００２３】この出願の請求項６（又は請求項１８）に記載の発明は、請求項１（又は請求項１３）乃至請求項５（又は請求項１７）のいずれかに記載のテキスト検索結果表示方法（又は装置）において、前記前記分割ステップ（又は手段）は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を、ファジイ・クラスタリング法を用いて複数個のグループに分割する、ことを特徴とするものである。

【００２４】そして、この請求項６（又は請求項１８）に記載の発明によれば、ある検索式により探し出された文書集合に対して自動的にorverlapping方式で主題内容によるファジイ分類（主題分類）が行われるため、検索漏れの防止による検索精度の向上が期待できる。

【００２５】この出願の請求項７（又は請求項１９）に記載の発明は、請求項１（又は請求項１３）又は請求項２（又は請求項１４）に記載のテキスト検索結果表示方法（又は装置）において、前記生成ステップ（又は手段）にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を幾つかのキーワードの組により表すものである、ことを特徴とするものである。

【００２６】そして、この請求項７（又は請求項１９）に記載の発明によれば、当該グループの属性を幾つかのキーワードの組を通して直観的に把握することができる。

【００２７】この出願の請求項８（又は請求項２０）に記載の発明は、請求項１（又は請求項１３）又は請求項２（又は請求項１４）に記載のテキスト検索結果表示方法（又は装置）において、前記生成ステップ（又は手段）にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を短い文章により表すものであることを特徴とするものである。

【００２８】そして、この請求項８（又は請求項２０）に記載の発明によれば、当該グループの属性を短い文章を通して誰にでも判りやすく理解させることができる。

【００２９】この出願の請求項９（又は請求項２１）に記載の発明は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合の特徴行列に対してファジイ・クラスタリングを行い、各文　毎に各分類カテゴリーへの所属度を生成する所属度生成ステップ（又は手段）と、前記生成された所属度を用いて、各文書を１若しくは２以上の分類カテゴリーに割り付ける文書割り付けステップ（又は手段）と、前記複数個の分類カテゴリーの中で、テキスト表示対象となる分類カテゴリーを選択するための分類カテゴリー選択ステップ（又は手段）と、前記分類カテゴリー選択ステップ（又は手段）で選択された分類カテゴリー内のテキストをそのグループに対する適合度の順に内容表示する表示ステップ（又は手段）とを具備する、ことを特徴とするテキスト検索結果表示方法（又は装置）にある。

【００３０】そして、この請求項９（又は請求項２１）に記載の発明によれば、overlapping手法を用いて各文書を１若しくは２以上の分類カテゴリーに割り付け、その状態にて選択された分類カテゴリー内のテキストをそのグループに対する適合度の順に内容表示するため、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができる。

【００３１】この出願の請求項１０（又は請求項２２）に記載の発明は、前記請求項９（又は請求項２１）に記載の発明において、前記文書割り付けステップ（又は手段）は、各文書をその所属度の上位ｋ個の分類カテゴリーに割り付ける、ことを特徴とするものである。

【００３２】そして、この請求項１０（又は請求項２２）に記載の発明によれば、請求項９（又は請求項２１）に記載の発明の効果に加え、各分類カテゴリーにつ

いて常に所属度の高い順に一定個数の文　を表示させることができる。

【００３３】この出願の請求項１１（又は請求項２３）に記載の発明は、前記請求項９（又は請求項２１）に記載の発明において、前記文書割り付けステップは、各文　をある閾値α以上の所属度値を有する分類カテゴリーに割り付ける、ことを特徴とするものである。

【００３４】そして、この請求項１１（又は請求項２３）に記載の発明によれば、請求項９（又は請求項２１）に記載の発明の効果に加え、各分類カテゴリーについて常にある閾値α以上の所属度値を有する文書を表示させることができる。

【００３５】この出願の請求項１２（又は請求項２４）に記載の発明は、前記請求項９（又は請求項２１）に記載の発明において、前記文書割り付けステップは、各文　をカテゴリーの確率分布を考慮して分類カテゴリーに割り付ける、ことを特徴とするものである。

【００３６】そして、この請求項１２（又は請求項２４）に記載の発明によれば、請求項９（又は請求項２１）に記載の発明の効果に加え、各分類カテゴリーについてカテゴリーの確率分布を考慮して関連する文書を表示させることができる。

【００３７】

【発明の実施の形態】以下に、本発明方法及び装置の好適な実施の形態を添付図面を参照しながら詳細に説明する。

【００３８】先ず、本発明方法及び装置が適用されたテキスト検索装置の構成を図１の機能ブロック図により概念的に示す。同図において、１は検索開始時に入力されるべきオリジナル検索質問（Original Query）や検索絞り込み時に入力されるべきフィードバック検索質問（FeedBack Query）を入力するための検索質問入力部（Query Inputと記す）であり、具体的には、周知のように、マウスやキーボード等の操作部とそれらの信号を処理する入力用ソフトウェアにより構成される。

【００３９】２は検索対象となるテキスト集合に相当するテキスト（文書）データベース（Doc.DBと記す）であり、具体的には、ハードディスクや光ディスク等の大容量記憶媒体に記憶されたテキスト集合やインターネット上に存在するホームページ等のテキスト集合がこれに相当する。

【００４０】３はテキスト検索システムの中枢に位置する検索エンジン（Search Engineと記す）であり、具体的には、周知のように、前述の検索質問入力部１から入力されるオリジナル検索質問（Original Query）やフィードバック検索質問（FeedBack Query）を検索条件として所定のアルゴリズムに従って検索式を展開し、前述の文書データベース２から関連するテキスト集合を抽出するソフトウェアがこれに相当する。

【００４１】４はこのようにして検索エンジン（Search Engine）３により抽出された関連するテキスト集合（Relevant Doc. Retrievedと記す）であり、後述するように、このテキスト集合４が本発明における加工処理の対象となる。

【００４２】５は本発明の要部に相当する加工処理部であり、この加工処理部５はテキスト集合４を各テキストの主題分析結果を用いて自動的に複数個のグループに分割する分割手段に相当するファジイ集合化部（Fuzzy Clusteringと記す）５ａと、こうして得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する主題分類情報生成部（Subject Generationと記す）５ｂとを中心として構成されている。

【００４３】ファジイ集合化部（Fuzzy Clustering）５ａ及び主題分類情報生成部（Subject Generation）５ｂの作用を図２に概念的に示す。同図において、符号４で示される実線にて囲まれた領域は検索エンジン（Search Engine）３にて抽出されたテキスト集合（Relevant Doc. Retrieved）の全体を表す。

【００４４】同様にして、符号４ａ，４ｂ，４ｃで示される破線にて囲まれた３つの領域はファジイ集合化部（Fuzzy Clustering）５にて分割された３つのグループのそれぞれを表す。

【００４５】符号Ａで示される黒塗り三角印は、検索開始時に入力されるオリジナル検索質問（Original Query）を表す。符号Ｂで示される×入り丸印は、オリジナル検索質問（Original Query）Ａの入力により検索抽出されたテキスト集合４の各構成テキストのそれぞれを表す。

【００４６】符号Ｃａ，Ｃｂ，Ｃｃで示される３個の白抜き三角印は、グループ４ａ，４ｂ，４ｃの属性を表現する主題分類情報（Group Subject）を表す。尚、これらの主題分類情報Ｃａ，Ｃｂ，Ｃｃは検索絞込みのために用いれば、フィードバック検索質問（FeedBack Query）としても好適なものである。

【００４７】符号Ｄａ，Ｄｂ，Ｄｃで示される３個の黒塗り丸印は、グループ４ａ，４ｂ，４ｃの重心を表す。同様にして、符号Ｄで示される黒塗り四角印は、テキスト集合４の重心を表す。

【００４８】図２から明らかなように、ファジイ集合化部（Fuzzy Clustering）５ａは、検索の結果得られたテキスト集合４に対して、公知のファジイクラスタリング処理を施すことにより、テキスト集合４を複数個（この例では３個）のグループ４ａ，４ｂ，４ｃに分割する。一方、主題分類情報生成部（Subject Generation）５ｂは、こうして得られた各グループ４ａ，４ｂ，４ｃのそれぞれについて、当該グループの属性を表現する主題分類情報Ｃａ，Ｃｂ，Ｃｃを生成する。図から明らかなように、このようにして得られる当該グループの属性を表現する主題分類情報Ｃａ，Ｃｂ，Ｃｃは、各グループ４ａ，４ｂ，４ｃの重心Ｄａ，Ｄｂ，Ｄｃとは異なるものであり、まさしくそれぞれのグループの属性を端的に表したものとなる。尚、これらのファジイ集合化部（Fuzzy Clustering）５ａ及び主題分類情報生成部（Subject Generation）５ｂの処理内容については、後に、更に詳しく説明する。

【００４９】図１に戻って、６は同様に本発明の要部に相当する検索結果提示部（Hit Windowと記す）であり、この検索結果提示部（Hit Window）６では、前述の経過により得られた情報（テキスト集合Ｂ、重心Ｄ、主題分類情報Ｃ）を所定の表示態様に加工したのち、ユーザ（Userと記す）７に対して提示する。それらの表示態様についても、後に詳細に説明する。

【００５０】尚、図１においては、実線により表された本発明による情報の流れと破線により表された従来装置による情報の流れとが同時に示されている。すなわち、従来装置にあっては、検索結果提示部（Hit Window）６では、破線矢印８に示されるように、検索の結果得られたテキスト集合Ｂをそのままユーザ７に提示するのみであり、この場合、テキスト集合Ｂに含まれるテキスト数が多量の場合、目的とするテキストを探し出すのにユーザは不便を来す。これに対して、本発明にあっては、検索結果提示部（Hit Window）６では、実線矢印９に示されるように、検索の結果得られたテキスト集合Ｂのみならず、各分類の重心（Clustercentroids）Ｄ並びに主題分類情報（Group Subject）Ｃまでもがユーザ７に提示されることとなるため、特に、この主題分類情報（Group Subject）Ｃを手掛かりとして、目的とするテキストを容易に探し出すことが可能となる。すなわち、実線矢印１０に示されるように、このようにして得られた主題分類情報Ｃ（図２のＣ１，Ｃ２，Ｃ３に相当する）をそのままフィードバック検索質問（FeedBack Query）Ｃとして検索質問入力部（Query Input）１に与えれば（図２の実線矢印１１に検索質問が分岐“Query Splitting”する様子を示す）、テキスト集合４を的確に絞込み、目的とするテキストを容易に探し出すことができ、すなわち高度な適応検索（relevance feedback）を行わせることができるのである。

【００５１】次に、以上概念的に説明したテキスト検索装置を、さらにその画面表示態様及びそれを実現するためのデータ処理を中心として、図３以下の図面を参照して詳細に説明する。

【００５２】本発明に係るテキスト検索装置におけるデータ処理の全体を図３のゼネラルフローチャートに示す。尚、このゼネラルフローチャートに示される処理は、所定のシステムメニューにおいて、そのメニュー項目のひとつを選択することにより起動される。

【００５３】同図において処理が開始されると、検索装置を構成する画像表示器の画面上には所定の表示態様により検索画面が表示される（ステップ３０１）。このようにして表示される検索画面の一例を図４に示す。同図に示されるように、表示画面は縦長長方形状のウィンドウＷ１により構成されており、その上部略３分の１の部分は検索質問入力領域Ａ１とされており、また下部略３分の２の部分は検索結果出力領域Ａ２とされている。

【００５４】検索質問入力領域Ａ１内には検索質問入力用のウィンドウＷ２が設けされており、このウィンドウＷ２の上側には、入力ガイド文（Enter Query in plain English）１２が、またその右側には、前述した検索エンジン（Search Engine）３に対する起動指令を与えるための起動ボタン（図中ＯＫと記す）１３と、検索質問（Query）を取り消すための取り消しボタン（図中ＣＡＮＣＥＬと記す）１４と、システムに対して操作支援等を求めるためのヘルプボタン（図中ＨＥＬＰと記す）１５とが設けられている。

【００５５】検索結果出力領域Ａ２内には検索結果出力用のウィンドウＷ３が設けられており、このウィンドウＷ３の右側にはスクロールバー１６が設けられている。更に、この検索結果出力領域Ａ２の下側には、検索結果としてテキスト全文出力を要求するための全文要求ボタン（図中Ｆｕｌｌ　Ｔｅｘｔと記す）１７と、ＱＢＥボタン１８と、検索結果の分類化を要求するための分類化要求ボタン（図中Ｇｒｏｕｐｉｎｇと記す）１９と、検索結果としてテキスト抄録出力を要求するための抄録要求ボタン（図中Ｓｕｍｍａｒｉｚｅと記す）２０と、画面を検索結果初期出力状態に戻すための復帰ボタン（図中Ｂａｃｋと記す）２１とが設けられている。

【００５６】尚、以上の各種のボタン１３，１４，１５，１６，１７，１８，１９，２０，２１の操作は、カーソルを希望のボタンに移動させた後、マウスのクリック操作等にて行われることは言うまでもない。

【００５７】そして、入力ガイド文（Enter Query in plain English）１２に従って、キーボードから検索質問を自然語（特に、この例では英語）にて、例えば、“I wantto know Clinton's political condition.”の如くに入力すると、この入力された検索質問２２はウィンドウＷ２内に表示されることとなる。

【００５８】この状態において、起動ボタン（図中ＯＫと記す）１３が操作されると、図３に戻って、検索／表示処理が実行され、検索エンジン（Search Engine）３が起動されて、検索質問に関連するテキスト集合４が文書データベース２より抽出され、この抽出されたテキスト集合の各構成テキストは検索質問２２との適合度の高い順にソートされ、そのタイトル２３のみがウィンドウＷ３内に表示される（ステップ３０２）。また、各テキストのタイトル２３の先頭部分には、当該テキストの検索質問に対する適合度を三段階（『高』、『中』、『低』）に区分して表す適合度マーク２４ａ，２４ｂ，２４ｃが表示される。ここで、黒色塗り潰しの丸印にて示される適合度マーク２４ａは適合度『高』に、灰色塗り潰しの丸印にて示される適合度マーク２４ｂは適合度

『中』に、白抜きの丸印にて示される適合度マーク２４ｃは適合度『低』にそれぞれ対応している。

【００５９】以後、図３に戻って、システム側においては文書処理機能の選択を待機する状態となる（ステップ３０３）。この状態において、図４の画面に示される分類化要求ボタン（Ｇｒｏｕｐｉｎｇ）１９が操作されると、本発明の要部である分類化処理が実行される（ステップ３０６）。

【００６０】分類化処理の詳細を図５に示す。同図において処理が開始されると、所定の案内画面を提示することにより、分類グループ数ｇの指定を待機する状態となる（ステップ５０１）。この状態において、分類グループ数ｇの指定（この例では『５』）が完了すると、本発明の特徴部分である文　特徴量の抽出処理（ステップ５０２）、ファジイ・クラスタリング処理（Fuzzy Clustering と記す）（ステップ５０３）、及び主題分類情報の生成処理（ステップ５０４）が順に実行される。

【００６１】文書特徴量の抽出処理（ステップ５０２）では、次のようにして、文書抽象化と文書特徴ベクトルの生成が行われる。文書は重み付けられた語の集合（語を構成要素とするベクトル）によって表され、文書の集合は語を構成要素とする行列として表される。そのため、各文書の特徴となる単語（重要語）を自動的に切り出し、単語の種類を次元mとし、各要素が文書単位の単語の出現頻度に比例するようなベクトル表現Ｆｉを用いることによって、文書は数１の如くに抽象化される。

【００６２】

【数１】

$$Fi=(fi1, fi2, \ldots, fim)$$

Ｆｉ：文書ｉの特徴ベクトル
ｆｉｊ：単語ｊの文書ｉに対する重み
（頻度、或は他の評価値）

数１

文　ベクトル集合の例を表１に示す。この例では、文書集合の構成文書（Ｆ１，Ｆ２，Ｆ３…）のそれぞれに含まれる重要語（Clinton, Singapore, China…）の重み（例えば、頻度）が示されている。

【００６３】

【表１】

	Clinton	Singapore	China	…
F1	0.8	0.4	0.0	
F2	0.6	0.7	0.0	
F3	0.5	0.0	0.7	
…				

文書ベクトル集合の例

表１

表１に示される文書ベクトル集合を文書空間に展開した例を図６に示す。この例では、前述の重要語（Clinton, Singapore, China…）を座標軸とする文書空間に文書集合の各構成文書（Ｆ１，Ｆ２，Ｆ３…）が展開されている。

【００６４】続くファジイ・クラスタリング処理（ステップ５０３）では、検索結果としての文書集合の特徴行列に対し、公知のＦＣＭ法を用いてファジイ・クラスタリングを行うことにより、次の２種類の分類情報（Ｖｃ，Ｕｉ）が生成される。

【００６５】１）　各分類の代表文書特徴ベクトルＶｃ

【数２】

$$Vc=(vc1, vc2, \ldots, vcn)$$

Ｖｇ：分類グループｇの代表文書ベクトル
ｖｃｊ：単語ｊの分類グループｃの代表文書に対する重み

数２

２）　各文書の各分類カテゴリーへの所属度Ｕｉ

【数３】

$$Ui=(ui1, ui2, \ldots, uig)$$

Ｕｉ：文書の各分類グループへ所属度ベクトル
ｕｉｃ：文書ｉの分類グループｃへの所属度

数３

文書分類所属度の例を表２に示す。この例では、各文書の所属度（Ｕ１，Ｕ２，Ｕ３…）が各分類グループ（Ｇ１，Ｇ２，Ｇ３…）毎に示されている。

【００６６】

【表２】

	G1	G2	…	Gg
U1	0.7	0.2	…	
U2	0.8	0.1	…	
U3	0.3	0.6	…	
…				

文書分類所属度の例

表２

続く分類主題情報の生成処理（ステップ５０４）では、次の２種類の方式により、分類主題情報の生成が行われる。

【００６７】１）　キーワード方式

このキーワード方式は、各分類グループの主題を幾つかのキーワードの組み合わせにより表現する方式であり、その際に、キーワードの抽出には次の２種類の方式が考えられる。第１の方式は、該当分類の代表文書ベクトルＶｃにおける重みの高い要素の単語を順番にｋ個抽出してそれらの単語をそのグループの主題を表す情報として用いるものである。第２の方式は、該当分類の文書集合に対して所属度の高い順にｒ個の文書ベクトルを選出し、そのｒ個の文書ベクトル集合において出現文書数の高いものから順にｋ個の単語を抽出して、そのグループの主題情報を表す情報として用いるものである。

【００６８】２）　テキスト方式

このテキスト方式では、上記のキーワード方式で主題情報を生成するために選出されたｒ個の文書の先頭段落のテキスト（タイトルを含む）に対し、キーワード方式で得られたキーワード主題情報を利用して文単位で文字列照合によりそれらのキーワードを最も多く所有するテキストを抽出し、そのテキスト文をそのグループの主題情報として用いるものである。

【００６９】このようにして得られた各グループの主題情報、すなわち分類主題情報（前述のキーワード群又はタイトル文等）は、後述するように、所定の提示順番にてユーザに提示されることとなる。ここで、検索された文書ｉの検索質問に対する適合度をＲｉ、分類グループの検索式への適合度をＧＲｃとすると、両者間には数４の関係が成立する。

【００７０】

【数４】

$$GRc=\sum_{i}^{rc} Ri / rc$$

ｒｃ：グループｃに対して所属度の高い順に選出された文書数
Ｒｉ：グループｃに対して選出されたｒｃ個の文書集合内のｉ文書の適合度

数４

ここで、数４に示された、グループｃに対して所属度の高い順に選出された文書数ｒｃ（ｃ＝１，…，ｇ：ｇ：分類数）の求め方を図７のフローチャートに示す。同図において、処理が開始されると、ｒｃの初期化（ｒｃ＝０）を行ったのち（ステップ７０１）、文書ｉの所属度の行データＵｉに対して最大の所属度が求められ（ステップ７０２）、その最大値と対応しているグループｃのメンバ数ｒｃが加算され（ステップ７０３）、以上の処理（ステップ７０２，７０３）がｉを＋１づつ加算しつつ（ステップ７０４）、その加算値がｉ＝ｎ（文書数）となるまで（ステップ７０５ＹＥＳ）繰り返されて、その結果ｒｃの値が最終的に求められることとなる。

【００７１】このようにして、分類主題情報の生成（提示順番の決定を含む）が完了すると（ステップ５０４）、求められた主題分類情報を用いた検索結果の動的表示処理が開始される（ステップ５０５）。

【００７２】検索結果の動的表示処理の詳細を図８のフローチャートに示す。同図において処理が開始されると、検索装置を構成する画像表示器の画面上に設定された検索結果出力領域Ａ２は、図９又は図１０に示されるように、上下に２分割され、これにより主題分類情報表示用ウィンドウ（Subject Window）Ｗ４と検索結果出力用ウィンドウ（Hit Window）Ｗ５とが現れる。そして、主題分類情報表示用ウィンドウ（Subject Window）Ｗ４において、所定の表示態様により、各分類主題情報の提示が行われる（ステップ８０１）。前述したように、この各分類主題情報の提示は、キーワード方式とテキスト方式とで行われる。

【００７３】キーワード方式による表示画面の一例を図９に示す。尚、この例では、検索されたテキスト集合が５個の分類グループに分割されている。同図に示される

ように、主題分類情報表示用ウィンドウ（Subject Window）Ｗ４内には、その左縁部に沿うようにして、分類グループ番号『１』～分類グループ番号『５』に対応する５個のグループボタン２５～２９が上下一列に配置されており、それらのグループボタン２５～２９の右側には、当該分類グループの主題を的確に表すキーワード群３０～３４が配列されている。この例では、分類グループ番号『１』に対応するグループボタン２５の右側には、キーワード群３０として、"SINGAPORE:CANE:PUNISH:US"が表示されており、分類グループ番号『２』に対応するグループボタン２６の右側には、キーワード群３１として、"DALAILAMA:MEET:CHINA:TIBET"が表示されており、分類グループ番号『３』に対応するグループボタン２７の右側には、キーワード群３２として、"MEET:LEADER:GOVERNMENT:OFFICIAL"が表示されており、分類グループ番号『４』に対応するグループボタン２８の右側には、キーワード群３３として、"NIXON:NATION:SINGAPORE:DIRECTIVE"が表示されており、分類グループ番号『５』に対応するグループボタン２９の右側には、キーワード群３４として、"QUESTION:CHARACTER:PEOPLE:POLITICS"が表示されている。

【００７４】また、これらの主題分類情報は、先に求められた提示順番に従い、検索質問（Query）との適合度の高いものから順に配列されている。すなわち、この例では、分類グループ番号『１』にて象徴される主題が最も検索質問との適合度が高く、分類グループ番号『５』にて象徴される主題が最も検索質問との適合度が低いこととなる。従って、ユーザー７は主題分類情報表示用ウィンドウ（Subject Window）Ｗ４内の表示順番から、自分の探している情報に最も近い分類グループを容易に知ることができ、しかもそれぞれの内容を端的に表すキーワード群３０～３４の内容に基いて、各分類グループの主題を大まかに確認することができる。そして、後に詳しく説明するように、分類結果表示処理（ステップ８０２）を起動することにより、当初の検索質問に沿うようにして、検索絞り込みを効率よく行うことができる。

【００７５】テキスト方式による表示画面の一例を図１０に示す。尚、この例でも、検索されたテキスト集合が５個の分類グループに分割されている。図に示されるように、主題分類情報表示用ウィンドウ（Subject Window）Ｗ４内には、その左縁部に沿うようにして、分類グループ番号『１』～分類グループ番号『５』に対応する５個のグループボタン２５～２９が上下一列に配置されており、それらのグループボタン２５～２９の右側には、当該分類グループの主題を的確に表す短いテキスト文３５～３９が配列されている。この例では、分類グループ番号『１』に対応するグループボタン２５の右側には、テキスト文３５として、"Clinton Protest Singapore Caning. Mulls Response"が表示されており、分類グループ番号『２』に対応するグループボタン２６の右側には、テキスト文３６として、"Clinton Meets With Dalai Lama"が表示されており、分類グループ番号『３』に対応するグループボタン２７の右側には、テキスト文３７として、"IndianLeader Meet Clinton"が表示されており、分類グループ番号『４』に対応するグループボタン２８の右側には、テキスト文３８として、"Nixon Had LivingWill"が表示されており、分類グループ番号『５』に対応するグループボタン２９の右側には、テキスト文３９として、"Clinton News Conferens-Text"が表示されている。

【００７６】また、これらの主題分類情報についても、先に求められた提示順番に従い、検索質問（Query）との適合度の高いものから順に配列されている。すなわち、この例では、分類グループ番号『１』にて象徴される分類グループの主題が最も検索質問との適合度が高く、分類グループ番号『５』にて象徴される分類グループの主題が最も検索質問との適合度が低いこととなる。従って、ユーザー７は主題分類情報表示用ウィンドウ（Subject Window）Ｗ４内の表示順番から、自分の探している情報に最も近い分類グループを容易に知ることができ、しかもそれぞれの内容を端的に表すテキスト文３５～３９の内容に基いて、各分類グループの主題を大まかに確認することができる。そして、後に詳しく説明するように、分類結果表示処理（ステップ８０２）を起動することにより、当初の検索質問に沿うようにして、検索絞り込みを効率よく行うことができる。

【００７７】次に、先に説明したファジイ・クラスタリングにより得られた各文書の各分類グループへの所属度Ｕｉを用いた、検索結果の最終表示のための処理について詳細に説明する。尚、この例では、分類結果の最終表示のためには３種類の処理が用意されており、これらの処理は図９又は図１０に示される画面において、グループボタン２５～２９のいずれか一つを操作することにより起動される（ステップ８０２）。

【００７８】先に説明したように、本発明では検索結果としての文書集合の特徴行列に対し、ＦＣＭ法を用いてファジイ・クラスタリングを行い、それにより各文書の各分類カテゴリーへの所属度Ｕｉが求められている。今仮に、５個の文書（００１，００２，００３，００４，００５）が存在し、それらの文書のそれぞれについて３個の分類カテゴリ（カテゴリ１、カテゴリ２、カテゴリ３）のそれぞれに対する所属度が表３の通りであると想定する。

【００７９】

【表３】

文書番号	カテゴリ１	カテゴリ２	カテゴリ３
００１	０．５０	０．３０	０．２０
００２	０．６０	０．１０	０．３０
００３	０．１０	０．８０	０．１０
００４	０．２５	０．３４	０．４１
００５	０．３５	０．１０	０．５５

割り付けの説明のための数値例

表３

以上の前提の元に、ファジイ分類結果の３種類の表示処理（１）～（３）を説明する。

【００８０】（１）　各文書の所属度の上位ｋ個の分類カテゴリーへ割り付ける場合

この表示処理にあっては、各文書（００１～００５）は所属度の高いものから順に選ばれたｋ個の分類カテゴリに割り当てられる。例えば、ｋ＝１とすると（２値化方式）、文書（００１）については最大所属度０．５０であるカテゴリ１に、文書（００２）については最大所属度０．６０であるカテゴリ１に、文書（００３）については最大所属度０．８０であるカテゴリ２に、文書（００４）については最大所属度０．４１であるカテゴリ３に、文書（００５）については最大所属度０．５５であるカテゴリ３にそれぞれ割り付けられる。これを分類カテゴリ（Ｇ１，Ｇ２，Ｇ３）別に整理すると、

カテゴリＧ１＝（００１，００２）　；Ｎ１＝２
カテゴリＧ２＝（００３）　；Ｎ２＝１
カテゴリＧ３＝（００４，００５）　；Ｎ３＝２

となり、分類グループＧ１に含まれる文書数Ｎ１は２個、分類グループＧ２に含まれる文書数Ｎ２は１個、分類グループ３に含まれる文書数Ｎ３は２個とされる。そして、このようにして各カテゴリに属することとされた文書が、後に詳細に説明するように、グループ番号の指定と共に検索結果出力用ウィンドウ（HitWindow）Ｗ５内に表示されることとなる。

【００８１】以上の表示処理（１）を実現するためのプログラムの一例を図１１に示す。同図において処理が開始されると、ｋ値の設定処理（ステップ１１０１）及びｉ，ｃ，Ｎｃの初期化処理（ステップ１１０２）を実行した後、文書ｉの所属度行データｉに対するソート処理（ステップ１１０３）、最大所属度データ値から順にｋ個のグループ番号を抽出する処理（ステップ１１０４）、及び該当するｋ個のグループに文書ｉを登録すると同時にメンバ数を加算する処理（ステップ１１０５）が、文書番号ｉがｎになるまで繰り返され（ステップ１１０６）、文書番号ｉがｎに達すると各グループ毎の文書割り付け結果を出力して処理が終了（ステップ１１０７）する。

【００８２】（２）　ある閾値α以上の所属度値を有する分類カテゴリに割り付ける場合

この表示処理にあっては、各文書（００１～００５）はある閾値α以上の所属度値を有する分類カテゴリに割り付けられる。ここで、αとしては、例えば１／ｇ（ｇ：分類数）とすることが考えられる。表３に示される例では、ｇ＝３、α＝０．３３となるため、文書（００１）については所属度値が０．３３以上であるカテゴリ１に、文書（００２）については同様な理由でカテゴリ１に、文書（００３）については同様な理由でカテゴリ２に、文書（００４）については同様な理由でカテゴリ２とカテゴリ３に、文書（００５）については同様な理由でカテゴリ１とカテゴリ３に割り付けられる。これを分類カテゴリ（Ｇ１，Ｇ２，Ｇ３）別に整理すると、

カテゴリＧ１＝（００１，００２，００５）　；Ｎ１＝３
カテゴリＧ２＝（００３，００４）　；Ｎ２＝２
カテゴリＧ３＝（００４，００５）　；Ｎ３＝２

となり、分類グループＧ１に含まれる文書数Ｎ１は３個、分類グループＧ２に含まれる文書数Ｎ２は２個、分類グループ３に含まれる文書数Ｎ３は２個とされる。そして、このようにして各カテゴリに属することとされた文書が、後に詳細に説明するように、グループ番号の指定と共に検索結果出力用ウィンドウ（HitWindow）Ｗ５内に表示されることとなる。

【００８３】以上の表示処理（２）を実現するためのプログラムの一例を図１２に示す。同図において処理が開始されると、α値の設定処理（ステップ１２０１）及びｉ，ｃ，Ｎｃの初期化処理（ステップ１２０２）を実行した後、文書ｉの所属度行データｉに対するｕｉｃ＞αのグループ番号を抽出する処理（ステップ１２０３）、該当する各グループに文書ｉを登録すると同時にメンバ数を加算する処理（ステップ１２０４）が、文書番号ｉがｎになるまで繰り返され（ステップ１２０５）、文番号ｉがｎに達すると各グループ毎の文書割り付け結果を出力して処理が終了（ステップ１２０６）する。

【００８４】（３）　カテゴリの確率分布を考慮して分類カテゴリに割り付ける場合

この表示処理にあっては、各文書（００１～００５）はカテゴリの確率分布を考慮して分類カテゴリに割り付けられる。ここで、文書の分類カテゴリーの確率分布（Ｐｃ）は数５に従って求められ、また分類ｃの文書数Ｎｃ

は数6に従って求められる。

【0085】

【数5】

$$Pc = rc / n$$

rc：（1）の2値化方式により得られた分類cの文書数

n　：文書集合の全文書数

【数6】

$$Nc = N(\alpha) \cdot Pc$$

$N(\alpha)$：（2）の割り付け方式により得られた各分類の文書表示数の和

表3に示される例では、P1＝0．4、P2＝0．2、P3＝0．4となり、またN（0．33）＝7となるため、N1＝2．8（約3）、N2＝1．4（約1）、N3＝2．8（約3）となる。これを分類カテゴリ（G1，G2，G3）別に整理すると、

カテゴリG1＝（001，002，005）　；N1＝2

カテゴリG2＝（003）　；N2＝1

カテゴリG3＝（002，004，005）　；N3＝2

となる。そして、このようにして各カテゴリに属することとされた文書が、後に詳細に説明するように、グループ番号の指定と共に検索結果出力用ウィンドウ（Hit Window）W5内に表示されることとなる。

【0086】以上の表示処理（3）を実現するためのプログラムの一例を図13に示す。同図において処理が開始されると、α値の設定処理（ステップ1301）、i，c，Ncの初期化処理（ステップ1302）、文書の分類カテゴリーの確率分布（Pc＝rc／n）を求める処理（ステップ1303）、分類cの文書数のNcを求める処理（ステップ1304）が順次に実行される。その後、文書cの所属度列データuicに対するソート処理（ステップ1305）、最大所属度値から順に対応のNc個のメンバの文書番号を抽出する処理（ステップ1306）、及び該当のグループcにNc個の文書を登録する処理（ステップ1307）が、分類cが分類数gになるまで繰り返され（ステップ1308NO）、分類cが分類数gに達すると（ステップ1308YES）、各グループ　の文書割り付け結果を出力して処理が終了する（ステップ1309）。

【0087】次に、以上説明した3種類の割り付け処理（1）～（3）のいずれかにて各分類グループに割り付けられた文　が、表示画面上の検索結果出力用ウィンドウ（Hit Window）W5内にどのような態様で表示されるかを説明する。

【0088】図9に示される画面上において、いずれかのグループボタン（この例では、グループボタン26）が指定操作されると、上述した3種類の割り付け処理（1）～（3）のいずれかにて各分類グループに割り付けられた文書に相当する短いテキスト文（この例ではタイトル等を含む当該テキストの先頭部分）40～44が、検索結果出力用ウィンドウ（Hit Window）W5内に表示されることとなる（ステップ802）。

【0089】すなわち、この例では、キーワード群31（"DALAILAMA:MEET:CHINA:TIBET"）にて象徴化される分類グループ番号『2』が指定されたことにより、検索結果出力用ウィンドウ（Hit Window）W5内には、これに関連する5個のテキスト文40（"Clinton Meets With Dalai Lama"）、テキスト文41（"Clinton,Gore MeetDalai Lama on Tibetan Right）、テキスト文42（"China Warns Clinton Notto Meet Dalai Lama"）、テキスト文43（"Clinton May Meet Dalai Lama before China Decision"）、テキスト文44（"Indian Leader Meet Clinton"）が表示されている。しかも、これらのテキスト文40～44は、図中『G』と記されたグループ適合度順指定ボタン51が操作されていることから、当該指定された分類グループ番号『2』で象徴化される分類グループとの適合度の順に配列して表示されている。尚、符号45，46はそれぞれその左側に位置するウィンドウW4，W5のスクロールバー、49は分類グループ数の表示である。

【0090】更に、検索結果出力用ウィンドウ（Hit Window）W5内において、各テキスト文40～44のそれぞれの先頭部分には、各テキスト文40～44が当該分類グループに対して有する適合度を3段階に表す3種類の適合度マーク（47a，47b，47c）と、各テキスト文40～44が当該検索質問22に対して有する適合度を3段階に表す3種類の適合度マーク（48a，48b，48c）が表示されている。この例では、当該分類グループとの適合度を表す適合度マーク（47a，47b，47c）は基本形状が雪印であり、適合度『高』に相当する適合度マーク47aについてはその中心の小円形部分を黒色塗り潰しに、適合度『中』に相当する適合度マーク47bについてはその中心の小円形部分を灰色塗り潰しに、更に適合度『低』に相当する適合度マーク47cについてはその中心の小円形部分を白抜きとしている。また、当該検索質問との適合度を表す適合度マーク（48a，48b，48c）は基本形状が丸印であり、適合度『高』に相当する適合度マーク48aについては黒色塗り潰しに、適合度『中』に相当する適合度マーク48bについては灰色塗り潰しに、更に適合度『低』に相当する適合度マーク48cについては白抜きとしている。

【0091】従って、この検索結果出力用ウィンドウ（Hit Window）W5内の表示内容40～44により、ユーザー7は検索結果であるテキスト集合の中で分類グループ番号『2』のグループに属するテキスト集合を、適合度マーク（47a，47b，47c）を頼りとして、該分類グループ『2』との適合度の高いものから順に確認しつつ、目的とする情報を的確に見つけ出すことができる。加えて、適合度マーク（48a，48b，48c）を参照することにより、各テキスト文40～44と検索質問22との適合度も知ることができるため、双方のマーク47，48を参考として、一層確実な検索絞り込みを行うことができる。尚、図示されてはいないが、図中『R』と記された検索質問適合度順指定ボタン50が操作された場合には、図8において分類主題表示指定処理（ステップ804）が実行されて、各テキスト文40～44は当該検索質問22との適合度の順に配列されて表示されることとなる。従って、検索質問適合度順指定ボタン50とグループ適合度順指定ボタン51とのいずれを選択するかにより、各テキスト文40～44の配列を変更しつつ、検索結果を所望の検索方向に沿って確認することができる。

【0092】一方、例えば図9に示される検索結果が表示されている状態において、操作支援要求ボタン（HELP）15が操作されると、図8に戻って、主題表示オプション処理（ステップ805）が実行され、主題分類情報表示用ウィンドウ（Subject Window）W4内の表示は、図10に示されるように、前述のキーワード方式からテキスト方式へと切り替わる。そのため、キーワード方式では当該分類グループの内容が把握しにくい場合でも、このテキスト方式による主題分類情報の表示によれば、当該分類グループにて象徴化される主題をより的確に知ることができる。尚、各ウィンドウW4，W5内に表示データが収まらない場合には、スクロールバー45，46の操作にて表示内容をスクロールしつつ確認できることは言うまでもない。

【0093】

【発明の効果】以上の説明で明らかなように、この発明によれば、文書検索結果に対する確認を容易として、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができ、しかも、提示された主題情報がデータを如何に効率的に絞り込めるかの指針にもなり、この付加された応答情報を利用して高度な適応検索（Relevance Feedback）を行わせることができる。

【図面の簡単な説明】

【図1】本発明方法及び装置が適用されたテキスト検索装置の構成を概念的に示すブロック図である。

【図2】ファジイ集合化部（Fuzzy Clustering）及び主題分類情報生成部（Subject Generation）の作用を概念的に示す説明図である。

【図3】本発明に係るテキスト検索装置の動作の全体を概略的に示すゼネラルフローチャートである。

【図4】本発明に係るテキスト検索装置においてグループ化処理を伴わない検索動作を実行させた状態を示す画面説明図である。

【図5】本発明に係るテキスト検索装置における主題分類情報の生成処理を中心として示すフローチャートである。

【図6】本発明に係るテキスト検索装置における文書抽象化と文書ベクトルの生成を概念的に示す説明図である。

【図7】本発明に係るテキスト検索装置におけるグループcのメンバ数rcを求めるための処理を示すフローチャートである。

【図8】本発明に係るテキスト検索装置における主題分類情報による検索結果の動的処理を示すフローチャートである。

【図9】本発明に係るテキスト検索装置においてグループ化処理を伴なう検索動作をキーワード方式にて実行させた状態を示す画面説明図である。

【図10】本発明に係るテキスト検索装置においてグループ化処理を伴なう検索動作をテキスト方式にて実行させた状態を示す画面説明図である。

【図11】本発明に係るテキスト検索装置にて検索結果をグループ別に表示するにおいて、各文書の所属度の上位k個の分類カテゴリーへの割り付け処理を示すフローチャートである。

【図12】本発明に係るテキスト検索装置にて検索結果をグループ別に表示するにおいて、α値以上の所属度値をもつ分類カテゴリーへの割り付け処理を示すフローチャートである。

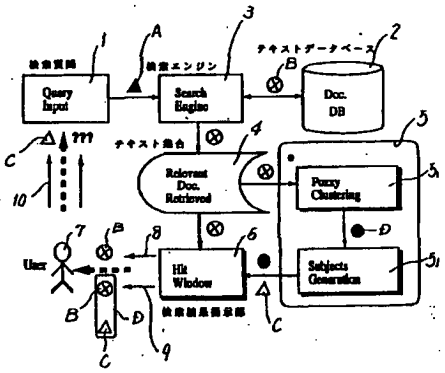
【図13】本発明に係るテキスト検索装置にて検索結果をグループ別に表示するにおいて、カテゴリーの確率分布を考慮した分類カテゴリーへの割り値付け処理を示す

フローチャートである。

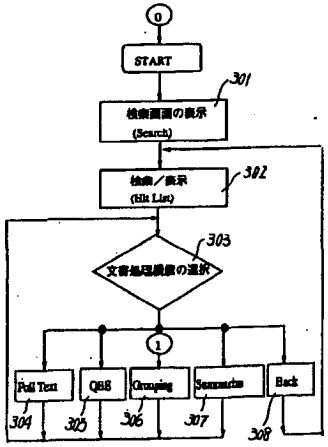
【符号の説明】

1　検索質問入力部
2　文　データベース
3　検索エンジン
4　抽出された関連テキスト集合
4a，4b，4c　分類グループ
5　加工処理部
5a　ファジイ集合化部
5b　主題分類情報生成部
6　検索結果提示部
7　ユーザー
12　入力ガイド文
13　起動ボタン
14　取り消しボタン
15　ヘルプボタン
16　スクロールバー
17　全文要求ボタン
18　QBEボタン
19　分類化要求ボタン
20　抄録要求ボタン
21　復帰ボタン
22　検索質問
23　テキスト集合を構成する各テキストのタイトル
24a，24b，24c　適合度マーク
25～29　グループボタン
30～34　キーワード群
35～39　テキスト文
40～44　テキスト文
45，46　スクロールバー
49　分類グループ数の表示
47a，47b，47c　グループ毎の適合度マーク
48a，48b，48c　検索質問に対する適合度マーク
49　分類グループ数の表示
50　検索質問適合度順指定ボタン
51　グループ適合度順指定ボタン
A　オリジナル検索質問
B　抽出された各構成テキスト
Ca，Cb，Cc　主題分類情報
Da，Db，Dc　グループの重心
A1　検索質問入力領域
A2　検索結果出力領域
W2　検索質問入力用のウィンドウ
W3　検索結果出力用のウィンドウ
W4　主題分類情報表示用ウィンドウ
W5　検索結果出力用ウィンドウ

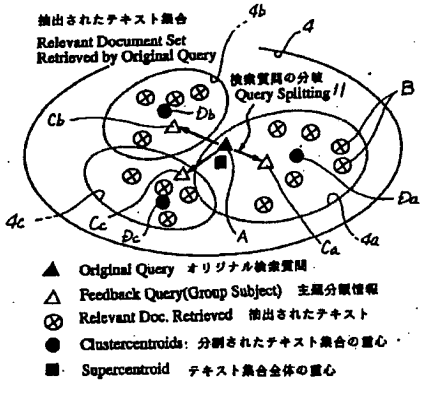
【図1】



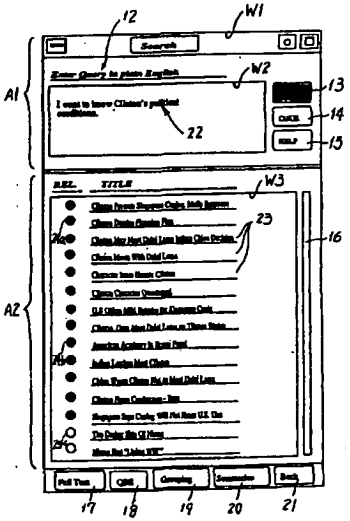
【図3】



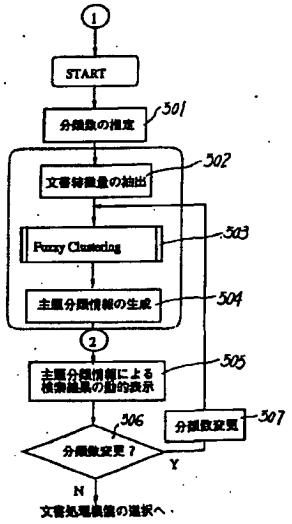
【図2】



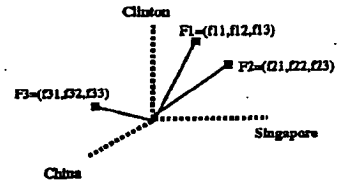
【図4】



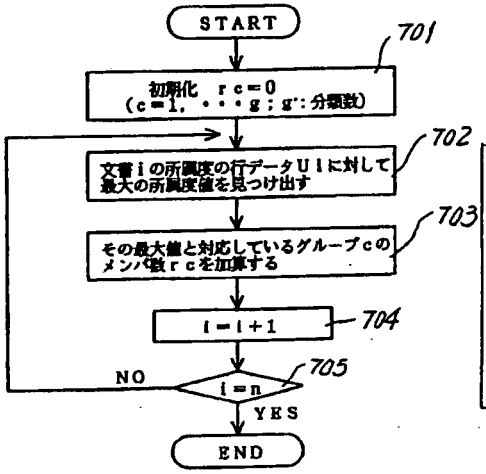
【図5】



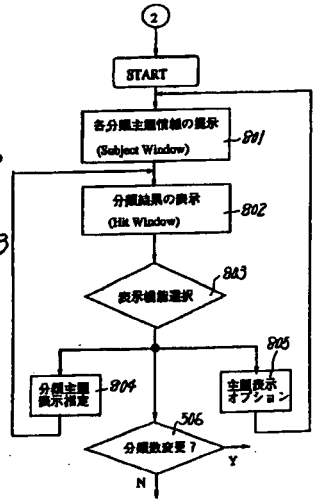
【図6】



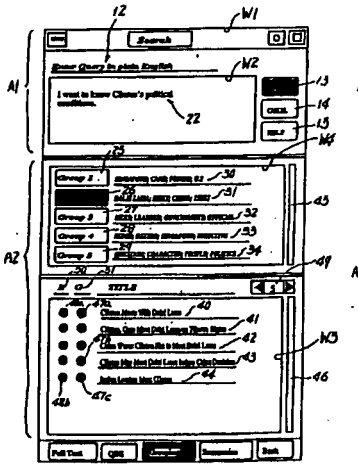
【図７】



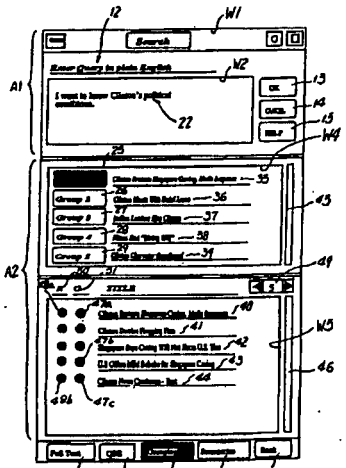
【図８】



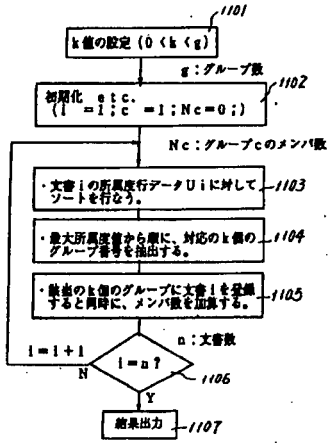
【図９】



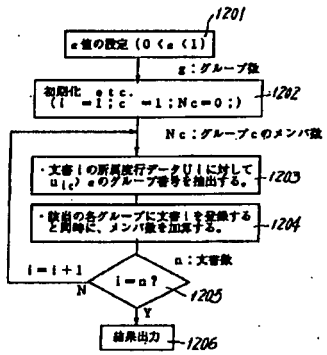
【図１０】



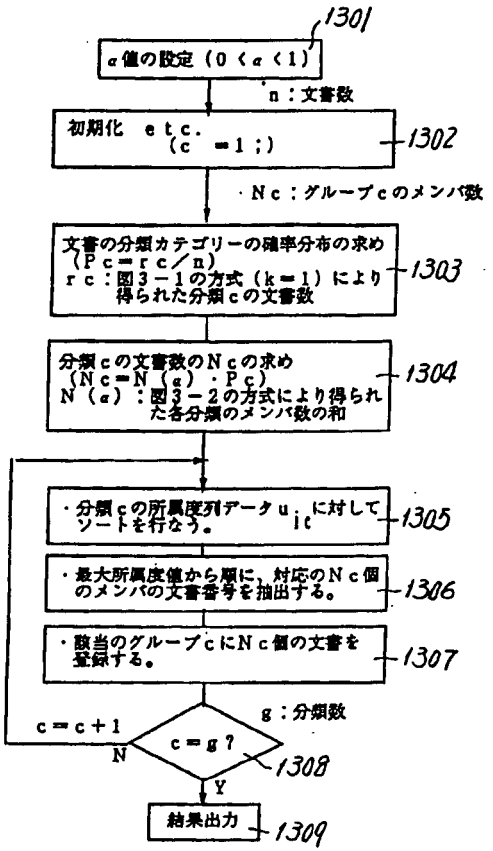
【図１１】



【図１２】



【図１３】


1301
α値の設定（0＜α＜1）
n：文書数
初期化　etc.
（c＝1；）
1302
Nc：グループcのメンバ数
文書の分類カテゴリーの確率分布の求め
（Pc＝rc／n）
rc：図３－１の方式（k＝1）により得られた分類cの文書数
1303
分類cの文書数のNcの求め
（Nc＝N（α）・Pc）
N（α）：図３－２の方式により得られた各分類のメンバ数の和
1304
・分類cの所属度列データu_{it}に対してソートを行なう。
1305
・最大所属度値から順に、対応のNc個のメンバの文書番号を抽出する。
1306
・該当のグループcにNc個の文書を登録する。
1307
g：分類数
c＝c＋1
N
c＝g？
1308
Y
結果出力
1309